Avaliação de Ferramentas de Apoio ao Ensino de Técnicas de Mineração de Dados em Cursos de Graduação

José Viterbo[1], Clodis Boscarioli[2], Flavia Cristina Bernardini[3], Mateus Felipe Teixeira[2]

[1]Instituto de Computação
Universidade Federal Fluminense (UFF)
Niterói – RJ – Brasil

[2]Colegiado de Ciência da Computação
Universidade Estadual do Oeste do Paraná (UNIOESTE)
Cascavel – PR – Brasil

[3]Instituto de Ciência e Tecnologia
Universidade Federal Fluminense
Rio das Ostras – RJ – Brasil

viterbo@ic.uff.br, clodis.boscarioli@unioeste.br,
fcbernardini@id.uff.br, mateusteixeira_@hotmail.com

Abstract. *The labor market requires more and more professionals with analytical sense to handle large data sets to produce useful knowledge for decision making. The data mining process (DM) is among the key concepts of knowledge discovery and involves the use of different machine learning algorithms applied to extract new standards in databases, commonly supported by tools that implement these different algorithms. The use of DM tools by undergraduate students is essential for them to acquire practical experience. It is done in courses such as Artificial Intelligence, Data Mining or other, addressing machine learning techniques. This study evaluates different tools used in DM teaching, in order to offer teachers a guide in the selection and use of these tools from the point of view of usability, measured by undergraduate students in the discovery and understanding of associated knowledge.*

Resumo. *O mercado de trabalho requer, cada vez mais, profissionais com senso analítico para lidar com volumosos conjuntos de dados de forma a produzir conhecimento útil à tomada de decisão. O processo de Mineração de Dados (MD) está entre os principais conceitos de descoberta de conhecimento e envolve o uso de diferentes algoritmos de aprendizado de máquina, aplicados para extrair novos padrões em bases de dados, comumente apoiado por ferramentas que implementam esses diferentes algoritmos. O uso de ferramentas para MD por alunos de graduação em disciplinas como Inteligência Artificial, Mineração de Dados ou outras que abordem técnicas de aprendizado de máquina, é fundamental para que eles adquiram experiência prática. Este trabalho avalia diferentes ferramentas utilizadas no ensino de MD, no intuito de oferecer aos docentes um guia na escolha e uso*

dessas ferramentas, do ponto de vista da usabilidade, aferida por alunos de graduação na descoberta e compreensão dos saberes associados.

1. Introdução

O mundo dos negócios vem passando por uma revolução impulsionada pelo uso analítico dos dados para orientar a tomada de decisões. O processo de análise eficiente de grandes massas de dados exige esforço multidisciplinar, uma vez que há desafios em diversas áreas que devem ser tratados. Neste cenário, surgem novas oportunidades de empregos, contudo, encontrar profissional qualificado à área de análise de dados passa a ser um grande desafio. A ciência de dados é um campo emergente e o cientista de dados é um profissional que precisa ter raciocínio lógico apurado, conhecimento profundo de estatística e modelagem matemática, além de domínio de sistemas computacionais e expertise do negócio e do mercado em que atuará [BRETERNITZ; SILVA 2013].

Dentre os conhecimentos fundamentais que esse profissional deve ter, se destacam diferentes tarefas e técnicas para apoiar as diversas etapas do processo de Mineração de Dados (MD) [REZENDE *et al*. 2005], [WITTEN 2011]. A MD é amparada por ferramentas que permitem a aplicação dos mais diversos algoritmos de aprendizado de máquina na análise de dados, sem que o usuário tenha que implementá-los. Nesse contexto, há a oportunidade de transformar o papel acadêmico de programador de algoritmos de mineração de dados para analista de inteligência de negócios, a partir do uso de ferramentas que deem suporte ao processo de descoberta de conhecimento em bases de dados, tornando-os capazes de analisar diferentes conjuntos de dados e de produzir relatórios que possam auxiliar na tomada de decisões.

Por outro lado, a MD apresenta novos desafios aos docentes que lecionam disciplinas nesta área. Um deles é a identificação de qual ferramenta computacional usar para introduzir o assunto em um curso de graduação. É importante que alunos de graduação em disciplinas como Inteligência Artificial, Aprendizado de Máquina, Mineração de Dados – ou outras cujo foco inclua suporte analítico para tomada de decisão, tomem contato prático com essas ferramentas, para vivenciar, de forma antecipada, processos de MD durante sua formação acadêmica.

Neste trabalho, investigamos como a usabilidade e a experiência do usuário com tais ferramentas pode melhorar a compreensão do processo de MD e do conhecimento descoberto e assim, melhorar a qualificação desse profissional. Foram avaliadas quatro ferramentas: Knime, Orange Canvas, Rapidminer Studio e Weka, e o resultado dessa avaliação pode auxiliar como um guia na escolha de ferramentas para ensino de MD.

A próxima seção traz os conceitos básicos relacionados à MD, de forma a contextualizar o leitor sobre o estudo realizado. Na Seção 3, são apresentados trabalhos relacionados a este estudo. Na Seção 4, são descritas as ferramentas avaliadas neste estudo. Na Seção 5, é discutida a metodologia empregada na pesquisa. Na Seção 6, é apresentada uma análise dos resultados obtidos com a avaliação das ferramentas de MD por estudantes de graduação em Computação. E, por fim, na Seção 7 são apresentadas as conclusões e perspectivas da pesquisa.

2. Mineração de Dados

O objetivo de um processo de MD é extrair conhecimento a partir de informações "ocultas" nos dados, que sejam úteis nas tomadas de decisões, utilizando métodos de diferentes áreas científicas, como Estatística, Banco de Dados, Inteligência Artificial, Aprendizado de Máquina e Reconhecimento de Padrões, caracterizando esse processo como interdisciplinar [WITTEN 2011]. As etapas desse processo são compostas pelo pré-processamento, construção de padrões e modelos e pós-processamento de dados, que envolvem, respectivamente, o tratamento dos dados de forma a qualificá-los para a entrada dos algoritmos de mineração de dados, a execução dos algoritmos para extração de padrões, e a compreensão das saídas para geração de conhecimento.

Na MD, a principal atividade – iterativa e interativa – envolve algoritmos de extração de padrões a partir da definição de uma tarefa de análise. As tarefas podem ser preditivas ou descritivas. As tarefas preditivas consistem na generalização de exemplos ou experiências passadas com respostas conhecidas, relacionadas às classes dos exemplos. Se a classe dos exemplos pertence a um conjunto discreto e finito de valores, a tarefa é denominada classificação; e se pertence a um conjunto numérico de valores, é denominada regressão. Já as tarefas descritivas consistem na identificação de comportamentos intrínsecos do conjunto de dados, sem que haja classes definidas dos exemplos. As tarefas descritivas podem ser de construção de regras de associação ou de agrupamento de dados [REZENDE *et al*. 2005], [WITTEN 2011].

Considerando que o processo de MD é interdisciplinar para compreender os algoritmos de aprendizado, seus ajustes de parâmetros e para avaliar os modelos construídos, ferramentas que possam auxiliar no entendimento de todo esse processo podem ser adotadas no processo de ensino e aprendizagem. Existem várias ferramentas disponíveis. Algumas delas são bastante difundidas e adotadas por docentes que ministram disciplinas correlatas a Mineração de Dados, por serem as mais antigas, a exemplo da Weka. Tal fato foi observado empiricamente pelos autores com diversos colegas docentes e observando ementas de diferentes disciplinas disponíveis na web. Essa adoção, porém, se dá mais por tradição que por uma decisão pautada em estudos de interação, preocupados com descobrir qual dessas ferramentas é a mais adequada para fins didáticos. O foco deste estudo parte da premissa que o uso de ferramentas de computação para ensinar disciplinas de computação deve ser pautado na avaliação de aspectos de interação com o usuário final, no caso em questão, aluno de graduação.

3. Trabalhos Relacionados

Vários são os autores que têm relatado a maneira que diferentes conteúdos de Inteligência Artificial e de Mineração de Dados têm sido abordados em disciplinas de graduação, bem como a importância dos conteúdos na formação. Em [FERREIRA *et al*. 2012] os autores analisam a abordagem de diferentes conceitos de Inteligência Artificial, incluindo conceitos aplicáveis à MD, apontando a importância em abordar tais conteúdos devido ao crescente uso de ferramentas de mineração de dados. Há uma análise da abordagem baseada em ferramentas para o ensino do processo de MD em [JAFAR 2010], que recomendam fortemente estender o currículo de cursos de Sistemas de Informação para incluir o estudo de MD em pelo menos uma disciplina. Para esses autores, as ferramentas permitem concentrar-se em ensinar os aspectos analíticos e o uso de algoritmos por meio de práticas, e concluem que os alunos ganham em

compreensão conceitual com essa abordagem. Em [MANASEER; MALIBARI 2012] também é reforçada a questão da prática, e apresentada uma proposta para o ensino de MD, permitindo aos alunos selecionar uma aplicação em dados reais. Esses autores afirmam ainda que a maioria da informação teórica ensinada é perdida após a graduação se não estiver relacionada a aspectos práticos e aplicações. As aulas práticas em cursos da área de Computação têm neste contexto, portanto, fundamental importância.

Em relação ao uso de ferramentas, em [KING e SATYANARAYANA 2013] são discutidas questões relevantes no ensino de Mineração de Dados em cursos de graduação, incluindo os principais tópicos que uma disciplina deve abordar. Os autores recomendam o uso de ferramentas de mineração de dados tradicionais (Weka, R e Rattle, sendo Rattle uma interface gráfica para uso da ferramenta R), por conterem uma ampla oferta de métodos para pré-processamento e extração de padrões. Em [CHAWLA 2005] é reportada a experiência do ensino de MD utilizando a ferramenta Weka, com destaque à importância da inserção dessa disciplina nos currículos de graduação. Os alunos do curso oferecido afirmaram que as aplicações práticas foram significativas no processo de aprendizado. Em [SATYANARAYANA 2013] o autor compara o uso de três ferramentas proprietárias – SAS, IBM SPSS Modeler e Matlab – e três ferramentas de código aberto – R, Weka e RapidMiner – no ensino de MD e conclue que, em cursos introdutórios de MD, o uso de ferramentas de código aberto é indicado, e em cursos avançados, o uso de ferramentas proprietárias pode ser também considerado.

É interessante observar que os diversos autores mencionam o uso de ferramentas, abertas ou não. No entanto, nenhum deles avalia, como neste trabalho, a questão da usabilidade dessas ferramentas no contexto da aprendizagem.

4. Ferramentas Avaliadas

A escolha das ferramentas analisadas, abaixo descritas, deu-se não apenas pelo fato de serem bastante conhecidas na área, mas também por terem uma variedade de funcionalidades em comum e possuírem versões de acesso livre, conforme já abordado em [BOSCARIOLI *et al*. 2015].

Knime [KNIME 2013]: ferramenta para uso na MD, Estatística e outras áreas. Possui diversos métodos que possibilitam extração de conhecimento de uma determinada base de dados. O funcionamento da Knime é todo baseado na ideia de adição de nodos de métodos a um fluxo de execução. A ferramenta apresenta algumas funcionalidades ainda não detalhadas, o que pode dificultar o seu uso.

Orange Canvas [CANVAS 2013]: ferramenta de código aberto de MD para classificação, regressão e tarefas descritivas de dados, e para mineração visual de dados. O fluxo de execução da ferramenta é dado por uma estrutura de nodos, em que cada nodo adicionado ao campo de fluxo executa uma determinada tarefa. Também apresenta um sistema de retorno ao usuário (*feedback*) para cada método, retornando-lhe os dados de entrada e de saída de cada método.

Rapidminer Studio [RAPIDMINER 2013]: ferramenta proprietária voltada à Estatística, Banco de Dados e processos de análises de dados que possui também uma versão para testes. Tem como foco disponibilizar ambiente de trabalho com presença de elementos gráficos que significam uma operação em questão, por exemplo, um método de MD. A ferramenta possui um fluxo de execução intuitivo e baseado em nodos que representam um determinado processo.

Weka [HALL *et al*. 2009]: ferramenta de aprendizado de máquina aberta usada para tarefas de MD, que contém técnicas para pré-processamento de dados, classificação, regressão, agrupamento, regras de associação e visualização.

5. Metodologia

Na avaliação das ferramentas, primeiramente realizou-se uma inspeção por meio do Percurso Cognitivo, um método de avaliação de IHC (Interação Humano-Computador) por inspeção cujo principal objetivo é avaliar a facilidade de aprendizado de um sistema interativo por meio da exploração de sua interface [WHARTON 1994]. Nesta abordagem, não há a necessidade da presença de usuários, e o(s) avaliador(es) devem associar objetivos específicos às ações e identificar problemas de interação de diferentes tipos de usuários e contextos reais.

Embora os percursos para realizar as ações sejam diferentes em cada ferramenta, pode-se concluir que o usuário vai tentar atingir o resultado correto desde que saiba o funcionamento teórico dos métodos implementados para a realização de ajustes dos parâmetros, bem como o conteúdo da base de dados a ser analisada para a verificação semântica dos resultados gerados. Os avaliadores deveriam identificar botões que possibilitam carregar a base de dados e especificar o caminho; botões que executam o algoritmo solicitado e configurar os parâmetros de entrada; e botões que possibilitam inserir métodos de visualização e, por fim, interpretar a saída dos algoritmos.

Há particularidades de interação entre essas ferramentas. Por exemplo, em todas as ferramentas o usuário necessita saber em que categoria um determinado item é classificado para poder encontrá-lo. No Weka, por exemplo, para a ação de abrir/carregar a base de dados podem acontecer erros na identificação do componente da ferramenta. Em alguns pontos, o componente chama-se "ARFF Reader" e noutros, é disponibilizado como "FILE", ficando implícito que este componente também poderá abrir/carregar um arquivo de extensão ".arff". Caso não saiba *a priori*, o usuário deverá aprender, enquanto realiza a tarefa, qual o fluxo necessário para realização das atividades de MD almejadas, em cada uma das ferramentas.

Compreendido o percurso cognitivo de uso, um teste de usabilidade foi então planejado e realizado com um grupo de 30 estudantes de Ciência da Computação da Universidade Federal Fluminense (UFF), aplicado imediatamente após terem cursado a disciplina de Aprendizado de Máquina, na qual foram ensinados os conceitos introdutórios de MD, suas tarefas e as principais técnicas de agrupamento e classificação de dados, usadas durante o teste de usabilidade com as ferramentas. Um estudo piloto foi previamente realizado em uma disciplina optativa chamada Introdução a Mineração de Dados no curso de Ciência da Computação da Universidade Estadual do Oeste do Paraná (UNIOESTE), campus de Cascavel.

Cabe destacar que, como parte das atividades dessa disciplina, estes alunos foram divididos em grupos de tal forma que cada 1/5 dos alunos ficou responsável pelo estudo prático da execução das tarefas de MD em uma das seguintes ferramentas: Knime, Orange Canvas, RapidMiner, Weka e CMSR Data Miner. Dessa maneira, cada uma das ferramentas avaliadas era conhecida por apenas 20% dos alunos (usuários), porém, todos deviam fazer as atividades em todas elas e, assim, responder a todas as questões propostas.

O teste de usabilidade consistiu na realização, em um laboratório de ensino de informática, de um conjunto de atividades propostas, relacionadas tanto a agrupamento quanto à classificação de dados, com tempo de duração suficiente para execução das tarefas e para responderem a um questionário contendo questões construídas pelos autores deste estudo sobre o perfil do usuário e a usabilidade da ferramenta.

Para o estudo das ferramentas com relação a agrupamento de dados, como exemplo, as seguintes tarefas deveriam ser realizadas pelos usuários: (1) Carregamento do arquivo de dados "Iris.arff"; (2) Execução do método *k-means* com $k = 3$ para o de agrupamento de dados (e demais parâmetros com valores *default*); (3) Interpretação dos resultados gerados observando todos os elementos de apresentação dos grupos na ferramenta; (4) Emprego de técnicas de visualização de dados sobre o resultado produzidos para observar se a compreensão dos resultados ficou mais fácil ou clara. Para a tarefa de Classificação, a variação desta metodologia está no item (2) onde o solicitado foi a execução do algoritmo *k-nn* (*k-nearest neighbors* vizinhos mais próximos), com parâmetros *default* das ferramentas, apenas variando o número de *k*.

6. Análise dos Resultados

A parte inicial do questionário procurou identificar aspectos relacionados ao perfil do usuário. A Figura 1 ilustra as respostas à Questão 1 – "Como você classifica seu conhecimento sobre Mineração de Dados?", à qual os alunos puderam escolher entre as opções "Básico/Introdutório", "Intermediário" ou "Avançado". Como era de se esperar, a maioria dos alunos se avaliou com nível de conhecimento "Básico/Introdutório", enquanto outros se classificaram em nível "Intermediário" ou "Avançado".

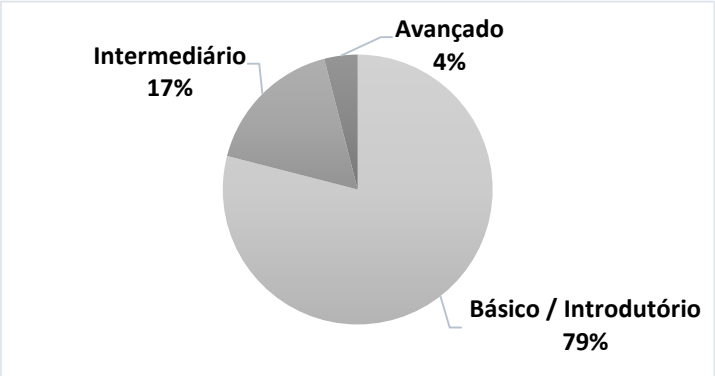

Figura 1. Nível de *expertise* de MD com os qual os estudantes se auto avaliam após terem cursado uma disciplina na área

A segunda parte do questionário procurou identificar extensivamente os diversos aspectos relacionados à usabilidade de cada uma das ferramentas. Neste trabalho, uma vez que o foco é a adequação das ferramentas para apoiar o ensino de MD, deu-se ênfase a questões com foco na facilidade de uso, no uso das técnicas de visualização para interpretação dos resultados e no quanto a navegação das ferramentas está relacionada aos conceitos da área estudada.

Os estudantes avaliaram (I) a facilidade da utilização de cada ferramenta respondendo à Questão 2 – "Esse sistema é fácil de utilizar?"; (II) a importância da visualização para interpretação dos resultados percebida em cada ferramenta respondendo às questões "As técnicas de visualização ajudaram na interpretação dos

resultados dos agrupamentos [Questão 3] ou da classificação [Questão 4]?"; (III) a disposição das tarefas na interface estaria coerente com os aspectos teóricos aprendidos, respondendo à Questão 5 – "A maneira em que os métodos aplicados para a execução das tarefas estão dispostos na interface da ferramenta permite concluir que esta execução segue, na ordem, as três principais etapas do processo de descoberta de conhecimento (pré-processamento, mineração de dados e pós-processamento)?"; e (IV) a categorização dos métodos disponíveis em cada ferramenta e a associação desses métodos às tarefas de mineração de dados, respondendo às Questões 6 – "Os menus e o esquema de navegação pela interface da ferramenta categorizam os métodos disponíveis de acordo com as tarefas mais conhecidas de mineração de dados?" e 7 – "A interface da ferramenta faz corretamente a associação dos métodos disponíveis com cada uma das tarefas de mineração de dados para as quais podem ser empregados?".

Para todas as perguntas, os usuários deviam atribuir notas de 1 a 5, onde 1 significava "Concordo Totalmente" e 5 significava "Discordo Totalmente". As médias dos resultados obtidos para a avaliação (I) – Questão 2 – para cada ferramenta constam na Figura 2. De acordo com os estudantes, a Orange Canvas é a que apresenta a interface onde é mais fácil encontrar as informações e elementos desejados, e Weka foi a pior avaliada, embora seja uma das ferramentas mais comumente adotadas. As médias dos resultados obtidos para a avaliação (II) – Questões 3 e 4 – para cada ferramenta podem ser observadas na Figura 3. Verifica-se que o Knime obteve uma avaliação melhor que o Orange Canvas, mas o Weka mais uma vez obteve um mau resultado.

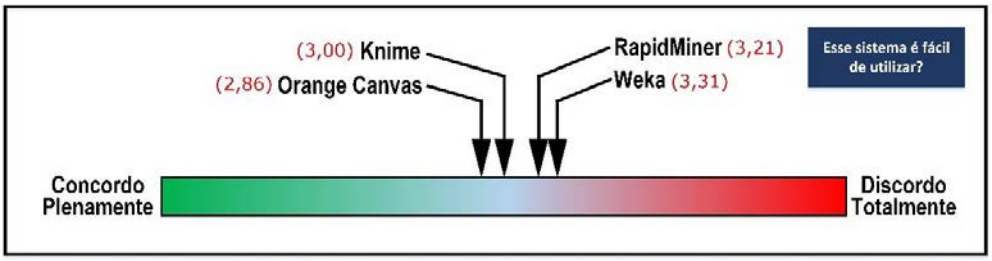

Figura 2. Avaliação dos estudantes em relação à facilidade de uso de cada ferramenta

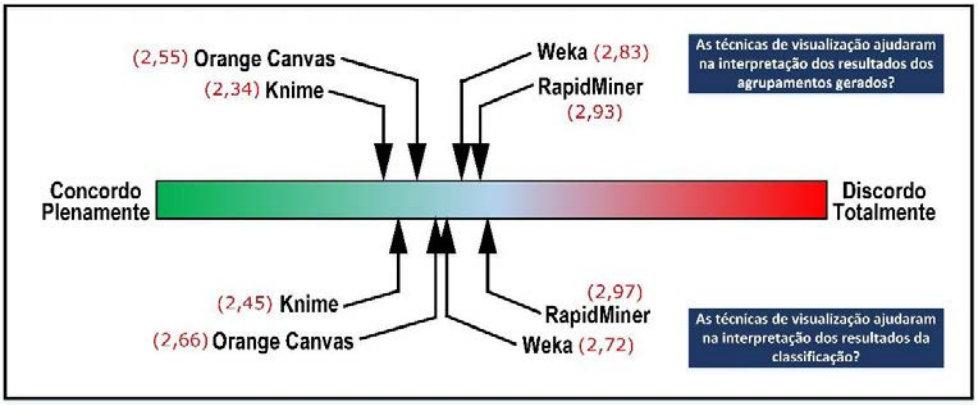

Figura 3. Avaliação dos acadêmicos em relação à importância da visualização para interpretação dos resultados

As médias dos resultados obtidos para a avaliação (III) – Questão 5 – para cada ferramenta podem ser observadas na Figura 4. A Knime se destacou mais neste aspecto,

seguida por RapidMinder. Weka novamente não foi bem avaliada. Já a ferramenta mais distante foi a Orange Canvas. No entanto, é curioso observar que essa ferramenta foi citada como a mais fácil de utilizar, como pode ser visualizado na Figura 2. Embora isso possa parecer contraditório – a disposição dos métodos na interface estar, dentre as ferramentas avaliadas, menos associada às três etapas do processo de MD e, ainda assim, ser a mais fácil de usar, compreende-se que, mesmo que os artefatos de interação não estejam à disposição como um fluxo de processo bem definido na interface, a forma como esses elementos estão dispostos ainda assim facilitam a interação do usuário, por prover-lhes subsídios de metacomunicação que os orientam na execução da tarefa de mineração escolhida. Por fim, as médias dos resultados obtidos para a avaliação (IV) – Questões 6 e 7 – para cada ferramenta podem ser observadas na Figura 5. Observa-se que em ambos os aspectos Knime obteve a melhor avaliação, seguida por Orange Canvas e Weka no primeiro aspecto e, Weka no segundo. Orange Canvas e RapidMiner obtiveram o pior resultado no primeiro aspecto, e RapidMiner, no segundo.

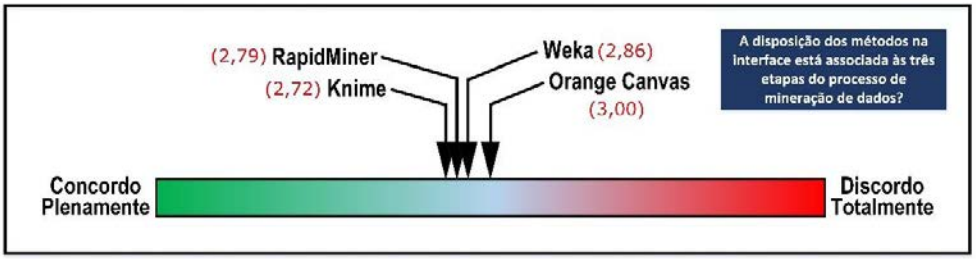

Figura 4. Avaliação dos estudantes em relação a se a disposição das tarefas na interface está coerente com os aspectos teóricos aprendidos

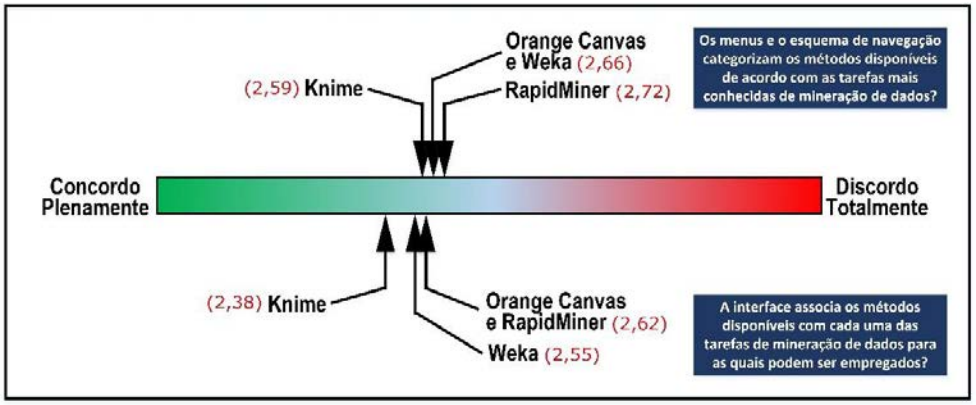

Figura 5. Avaliação dos estudantes em relação à categorização dos métodos e à associação desses às tarefas de mineração de dados

Muito embora todas as ferramentas sejam bastante maduras e abrangentes nas tarefas e técnicas de mineração de dados, nota-se que há diferenças significativas entre elas quando comparadas pelo prisma da interação com o usuário, o que pode ser bastante restritivo para usuários iniciantes, como os estudantes que as utilizam para consolidação de conhecimentos teóricos vistos em sala de aula.

7. Conclusões e Perspectivas

Neste trabalho foram avaliadas sob o ponto de vista de usabilidade, quatro ferramentas que apoiam o processo de MD, do ponto de vista de alunos que estão aprendendo o

referido processo. Pôde-se perceber que, em geral, as ferramentas analisadas dão forte ênfase à visualização de dados, permitindo ao usuário, por exemplo, visualizar os grupos gerados pelos métodos de agrupamento, favorecendo sua compreensão, interpretação e extração de conhecimento sobre o conjunto de padrões. Essa característica é de grande valia aos estudantes para aquisição/consolidação de conhecimentos na área, de forma aplicada e prática.

O uso de ferramentas de MD necessita, em primeira instância, de conhecimento técnico sobre o assunto. Quão maior for o domínio teórico e o conhecimento dos jargões técnicos, estima-se que menor será o tempo de aprendizado de uma ferramenta de MD. Porém, a interface desempenha papel fundamental para o domínio e produtividade em tais ferramentas, dado que influencia diretamente na usabilidade, em aspectos como experiência do usuário e memorização. Portanto, a interface dessas ferramentas passa a ter importância maior, em se tratando de primeiro uso, para fins de aprendizado e aplicação de MD em todos os seus conceitos e nuances.

De modo geral, a usabilidade e a facilidade de operação dos métodos de cada ferramenta são boas, o que as permite abranger públicos com qualquer nível de conhecimento elas, ou até mesmo sobre os métodos que disponibilizam, já que para a execução todas apresentam fluxos de execução que embora diferentes, são bastante claros como processo, tendo início, meio e fim.

A ferramenta Weka, uma das mais utilizadas em cursos de graduação e pós-graduação no Brasil, muito pelo seu pioneirismo, foi, no geral, a que obteve uma das piores avaliações pelos estudantes quanto à sua usabilidade. Outras ferramentas, menos adotadas, como Knime e Orange Canvas, tiveram melhores avaliações gerais. Esse resultado é bastante interessante para mostrar que a preocupação com a escolha de uma ferramenta, mesmo para o ensino de estudantes de cursos da área de Computação, deve estar também pautada em aspectos de interação que podem ser facilitadores ao processo de apropriação do conhecimento de MD.

A avaliação da usabilidade de ferramentas para MD pode aumentar o potencial dessas ferramentas, utilizando o conhecimento do usuário e ajudando-o a obter resultados cada vez mais expressivos das bases de dados. Por outro lado, a avaliação realizada mostra que outros aspectos de IHC necessitam ser abordados mais cuidadosamente para extrair diretrizes que orientem no *design* de interação de ferramentas para auxiliar o processo de MD.

Como trabalhos futuros os autores pretendem ampliar a avaliação ora apresentada: (i) incluindo novas tarefas de mineração de dados, abrangendo ao menos um método de cada uma das principais tarefas de MD (Classificação, Associação, Agrupamento e Regressão); e (ii) avançando na indicação de qual dessas ferramentas seria mais indicada para o ensino dessa disciplina, a partir da avaliação da interação dos estudantes. Com isso, há também a intenção de avaliar mais profundamente essas ferramentas e o quão experiente o usuário tem que ser para poder usar seus algoritmos. Finalmente, pretende-se propor uma ontologia que disponha de todos os métodos de MD, de forma que melhore a interação do usuário (estudante) com tais ferramentas.

Referências

BOSCARIOLI, C.; VITERBO, J.; TEIXEIRA, M. F.; ROHSIG, V. Analyzing HCI Issues in Data Clustering Tools. (2014) In: Proceedings of the 16th International Conference on Human-Computer Interaction. London: Springer, v. 8521. p. 22-33.

BRETERNITZ, V. J.; SILVA, L. A. "Big Data: Um Novo Conceito Gerando Oportunidades e Desafios" (2013), Revista Eletrônica de Tecnologia e Cultura – RETC. 13ª Ed., FATEC, ISSN: 2177-0425, p. 106-113.

CANVAS. Orange Canvas: Data Mining – Fruitful and Fun. (2013). Disponível em http://orange.biolab.si/. Acessado em 05 de março de 2016.

CHAWLA, N.V. Teaching Data Mining by Coalescing Theory and Applications. In: Proc. 35th ASEE/IEEE Frontiers in Education Conference.

FERREIRA, P. C.; BERNARDINI, F. C.; SILVA, C. R. C.; VITERBO, J. How Computer Science Undergraduate Programs in Brazilian Public Universities Address Artificial Intelligence and its Topics: A Preliminary Report (2012). Learning and Nonlinear Models, v. 12, pp. 73-85.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update (2009); SIGKDD Explorations, Volume 11, Issue 1.

JAFAR, J. M. A Tools-Based Approach to Teaching Data Mining Methods (2010). Journal of Information Technology Education: Innovations in Practice. Beachboard, J (Editor). Volume 9, p. 1-24.

KING, B.R.; SATYANARAYANA, A. Teaching Data Mining in The Era of Big Data (2013). In: Proc. 120th ASEE Annual Conference & Exposition.

KNIME. Knime: Open for Innovation (2013). Disponível em https://www.knime.org/. Acessado em 29 de maio de 2015.

MANASEER, S. A.; MALIBARI, A. Improve Teaching Method of Data Mining Course (2012). I. J. Modern Education and Computer Science, Volume 2, p. 15-22.

RAPIDMINER. Rapidminer: Analytics for Anyon (2013). Disponível em https://rapidminer.com/. Acessado em 08 de março de 2016.

REZENDE, R. O.; PUGLIESI, J. B.; MELANDA, E. A.; DE PAULA, M. F. (2005). Mineração de Dados. Cap 12. In: REZENDE, S. O. (Org.). Sistemas Inteligentes: Fundamentos e Aplicações. Barueri-SP: Ed. Manole, p. 307-335.

SATYANARAYANA, A. Software Tools for Teaching Undergraduate Data Mining Course (2013). In: Proc. ASEE MidAtlantic Conference.

WHARTON, C., RIEMAN, J., LEWIS, C., and POISON, P. The cognitive walkthrough method: A practitioner's guide (1994).

WITTEN, I.; FRANK, E.; HALL, M. Data Mining: Practical Machine Learning Tools and Techniques (2011). Morgan Kaufmann Publishers.